5-2014

Biophysical and Structural Characterization of Nuclear Hormone Receptors

Mark Remec
mremecpa@vols.utk.edu

Biophysical and Structural Characterization of Nuclear Hormone Receptors

Mark Remec Pavlin
College Scholars: Biophysics
Mentor: Elias J. Fernandez, BCMB
Spring 2014

Table of Contents

Introduction and Background:

Results and Conclusions:

All cellular organisms require the ability to integrate and respond to intra and extra-cellular signals. In eukaryotic organisms, complex signaling machineries have been developed for this purpose. The Nuclear Hormone Receptor (NR) superfamily of proteins is one such mechanism. This family encompasses a large number of structurally and functionally similar proteins that regulate a wide variety of physiological processes, such as toxin clearance, growth and development, and homeostasis, through activation of transcription of target genes. This thesis will deal primarily with biophysical, biochemical, and structural investigations into the Nuclear Receptor family of proteins, with the goal of furthering our understanding of the structures and functions of these proteins.

The discovery of hormones and steroids in the early 20th century catalyzed the investigations that would eventually lead to the discovery of Nuclear Receptors. Hormones and steroids were first analyzed for the large-scale effects that could be observed in organisms upon their addition; a seminal study involved attributing early amphibian metamorphosis to the addition of mammalian thyroid extracts. Further investigations revealed significant effects potentiated by other hormone and steroid molecules, including cortisone and estrogen. The biological significance of steroid and hormone molecules led to a massive effort to organically synthesize these molecules in the lab, which led to the famous race between Woodward and Robinson to achieve a total synthesis of Cholesterol. These successes furthered the pharmacological interest in the function of hormones and steroids, and by 1961 Jensen had radioactively tracked the hormone oestradiol-17β to the nucleus, in complex with a protein. This soon led to a new model for the action of hormones and steroids, whereby the function of the ligand was to interact with and thereby activate a receptor protein. Improvements in technology that spurred the molecular biology revolution also led to the cloning of the Glucocorticoid and

Thyroid Hormone Receptor in the lab of Ron Evans, and the Estrogen Receptor in the labs of Pierre Chambon and Geoffrey Greene in the mid 1980's, which led to a rapid increase in the understanding of NR structure and function. Since these seminal experiments, a great many NR receptors have been recombinantly purified and examined structurally and functionally, both individually and in various complexes with their physiological partners and ligands. Current work focuses on ever larger complexes with full-length receptors and their interacting partners to more comprehensively understand their physiological functions.

Though many questions remain, a great deal has been uncovered about the structure of NRs. It is now known that NRs are characteristically modular proteins with several domains. Most receptors contain six domains: a highly variable N-terminal domain containing an Activation Function 1 (AF-1) domain, a well-conserved zinc-finger DNA-binding Domain (DBD), a hinge region, a Ligand Binding Domain (LBD), and an Activation Function 2 (AF-2).

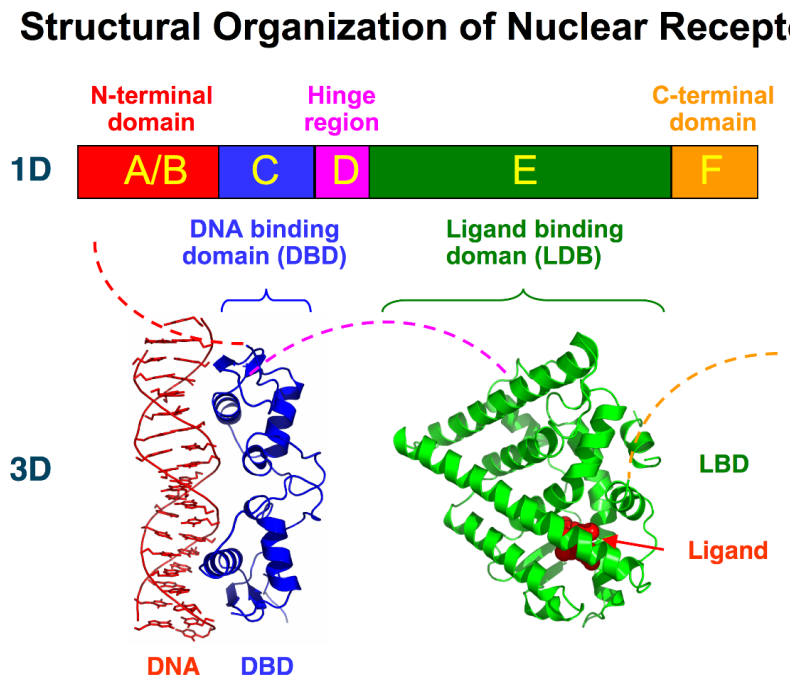

Fig 1.1: General structural organization of nuclear receptors. Source: http://upload.wikimedia.org/wikipedia/commons/3/3e/Nuclear_Receptor_Structure.png

The structures of the individual DBDs and LBDs have been studied intensively by x-ray crystallography and biochemical assays, though larger fragments and full-length studies of nuclear receptors have been limited by the difficulty of recombinant purification and structural characterization. Indeed, only recently have DBD-LBD complexes been successfully analyzed by crystallography, scattering, or cryo-EM. These analyses have revealed a general paradigm for the function of NRs: ligand activation by binding in the LBD causes dissociation of NRs from their co-repressor complexes, and subsequent translocation (if necessary) to their DNA promoter sequences as monomers, dimers or heterodimers, upon which the receptors bind coactivator molecules to initiate the assembly of the transcription machinery. The dimerization interface, which is usually composed of approximately three helices from each receptor subunit, mediates allosteric communication between the dimer subunits. The mechanism of coactivator binding is fairly well understood in individual cases. In general, ligand binding to the ligand-binding cavity induces a conformational change that pulls Helix 12 from a flexible conformation extending from the receptor surface to a static conformation lying along the receptor surface. This conformational change dissociates corepressor complexes and creates a coactivator interaction surface that consists of a hydrophobic groove with “charge clamp” residues to bind to specific residues in cognate coactivator proteins. The LBDs also contain the dimerization interface in most cases, whereby homodimerization or heterodimerization may occur.

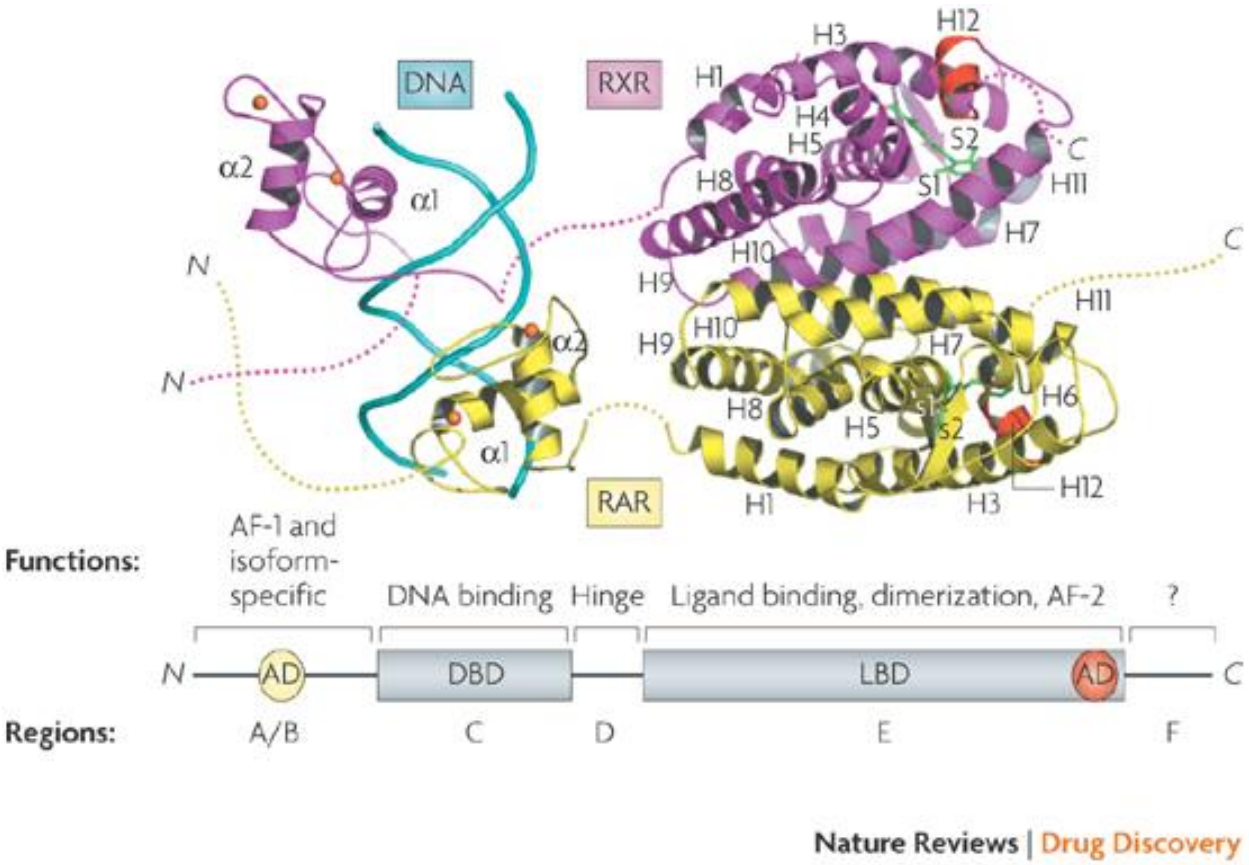

Fig 1.2: Diagram of helices in the DBD and LBD of a Retinoid X Receptor : Retinoic Acid Receptor heterodimer. The coactivator-interaction surface near H12 and the dimerization interface near H10 are clearly shown. Source: http://www.nature.com/nrd/journal/v6/n10/fig_tab/nrd2398_F1.html

The structure of the LBD, therefore, determines specificity for ligand, coactivator, and other NR interactions, and this structure is therefore relatively variable among the different members of the NR superfamily. The DBD is by comparison a more conserved subunit, and is most important for its zinc-finger DNA recognition function. The DBD connects to the LBD via a short, flexible hinge domain whose functional properties remain unclear, despite being evidently significant. Most NRs also contain a variably-size N-terminal domain with an AF-1. These domains are very weakly conserved throughout the superfamily, and their high degree of flexibility has precluded extensive study. While it is understood that this N-terminal domain is important in regulating transcription, mechanisms to explain its function are not yet available.

A detailed atomic view of both the LBD and DBD is necessary to understand much of

modern NR research. Many NR LBDs have now been crystallized, and all reveal a globular structure consisting of approximately 12 α-helices. These helices form a set of three antiparallel sheets, the interior of which contains the ligand-binding pocket. These ligand binding pockets determine their specificity primarily through hydrophobic interactions, though hydrogen bonding interactions and specific residue interactions are often critical in determining the orientation and affinity for the precise ligand required.

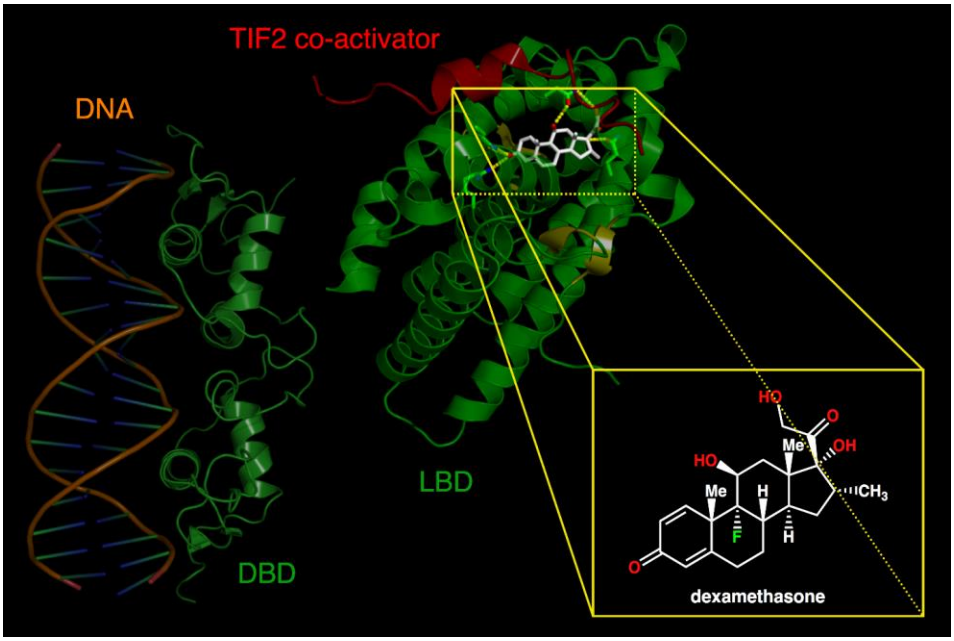

Fig 1.3: Glucocorticoid receptor with dexamethasone bound in the ligand binding pocket. Several charge clamps are indicated by yellow lines. Source:http://upload.wikimedia.org/wikipedia/commons/f/f0/Glucocorticoid_receptor.png

The AF-2 domain consists largely of Helix 12, though other helices are also involved. Upon binding of agonist ligand, Helix 12 rearranges to form a coactivator interaction surface. Coactivators with helical LXXLL sequences are able to bind to these surfaces primarily through hydrophobic interactions, with “charge clamp” interactions at each end of the helix mediated usually by glutamate and lysine residues on the NR. When bound to antagonist ligand or unliganded, most receptors either cannot form the coactivator interaction surface or have a corepressor outcompeting the coactivator for the surface. Therefore, ligand binding is a critical mediator of NR activity.

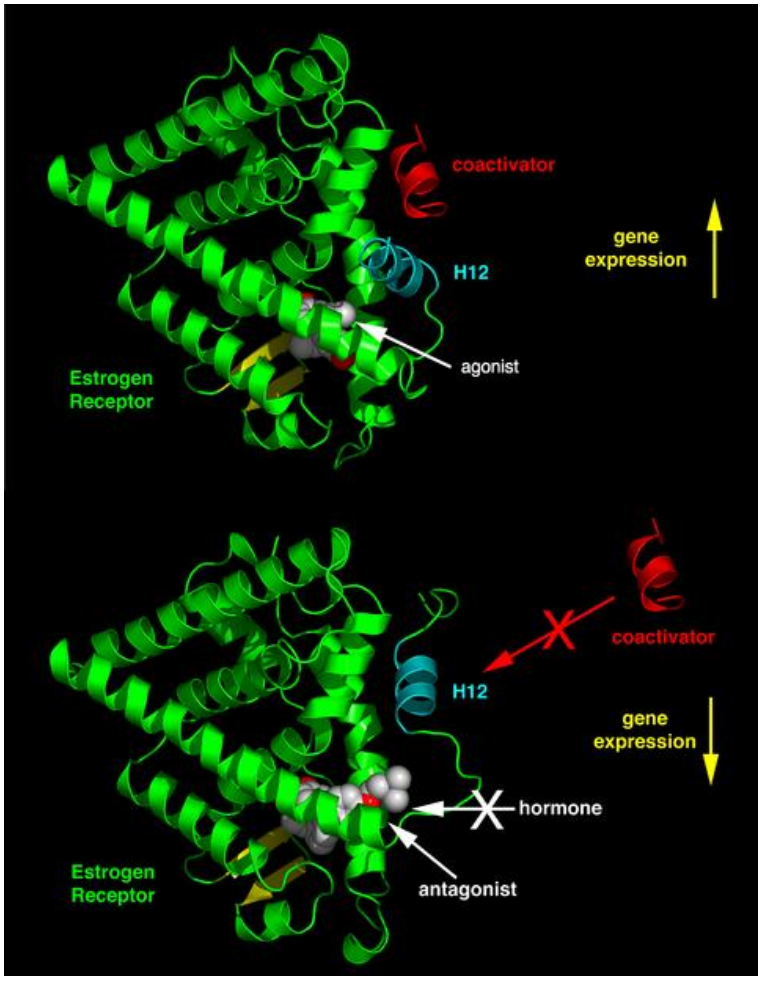

Fig 1.4: Conformation change associated with the binding of ligands and its physiological impact. Source:http://upload.wikimedia.org/wikipedia/commons/thumb/d/d4/NR_mechanism.png/480px-NR_mechanism.png

The DBD is a highly conserved subunit that consists of two zinc-finger structures. In each, a zinc atom is coordinated by four cysteine molecules. Two α-helices also contribute to DNA-binding specificity and overall protein stability. The DBD has also been observed to be a potent allosteric communicator to other parts of the NR, though precise mechanisms for allosteric communication, especially involving the hinge region between LBD and DBD, remain unclear. Finally, the DBD can in certain cases contribute a dimerization interface.

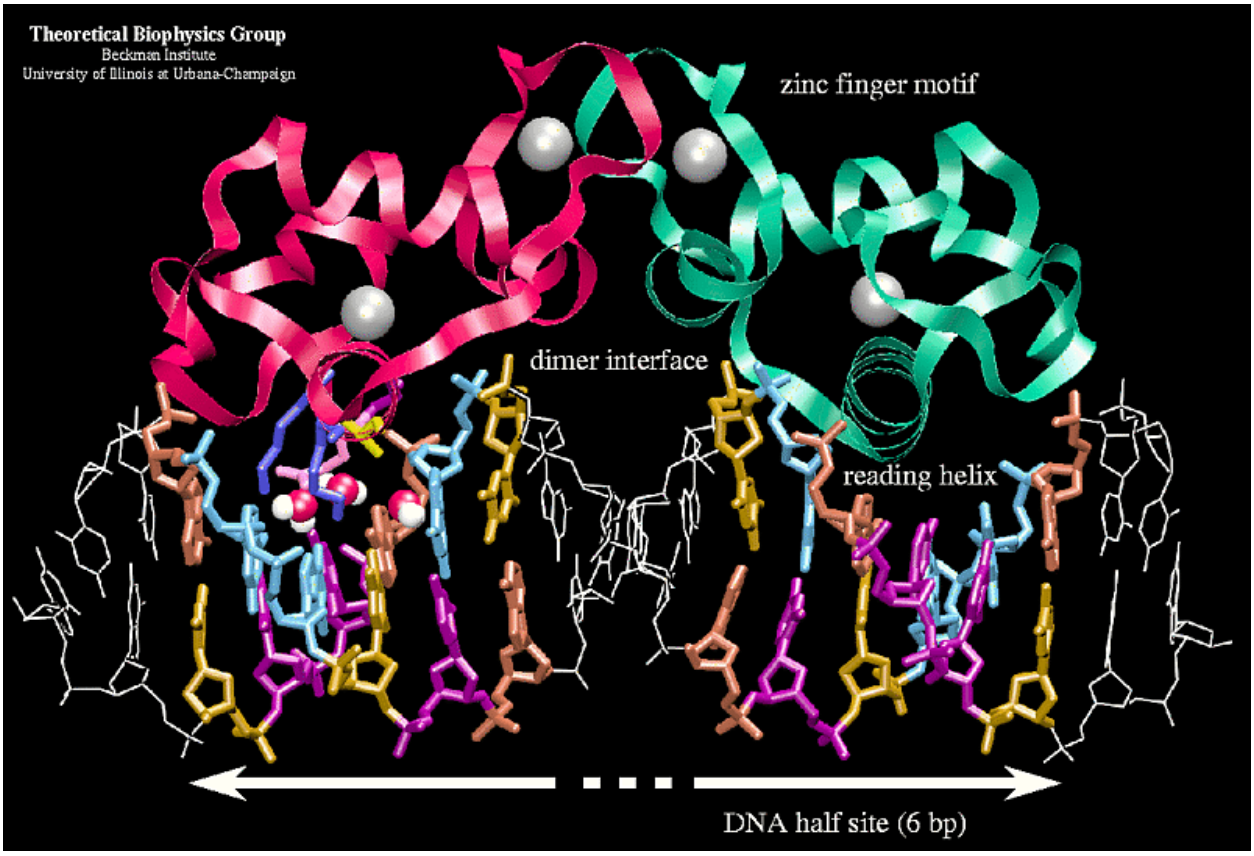

Fig 1.5: DBDs of an Estrogen receptor dimer on DNA, showing a dimerization interface in the DBD. Source: http://www.ks.uiuc.edu/Research/pro_DNA/ster_horm_rec/dbd/dbd_big.gif

The N-terminal regions beyond the DBD remain poorly understood. Atomic resolution of these regions has not yet been achieved, however several biochemical assays have yielded hints as to the structure and function of these domains. Several studies have shown these regions to be mostly disordered, with little secondary structure. However, it has been seen that these regions can adopt secondary structure under a variety of conditions, including DNA binding and interaction with other transcription factors. Furthermore, this region contains several phosphorylation regions that have been linked to its structural instability. Despite these insights, an integrated understanding of the many possible functions of these N-terminal regions is still not available. Similarly, the flexible hinge region between the LBD and DBD has also been

noted to have several important phosphorylation sites, though the functional mechanisms of the hinge remain unclear.

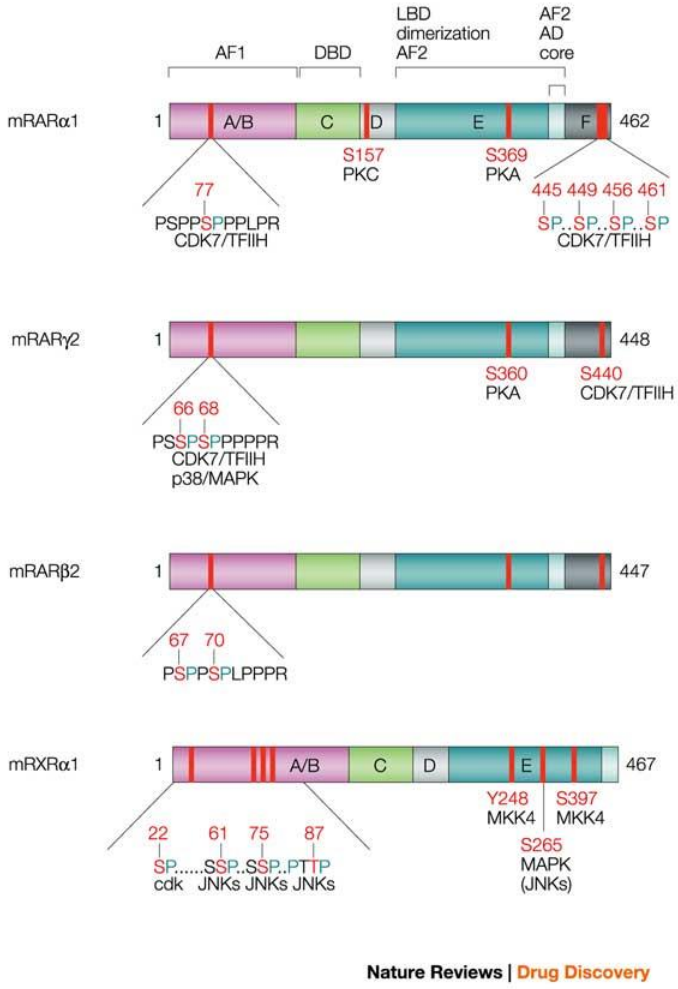

Fig 1.6: Diagram indicating phosphorylation sites of some isoforms of RAR and RXR. The N-terminus harbors many phosphorylation sites. Source:http://www.nature.com/nrd/journal/v3/n11/images/nrd1551-i1.jpg

There are many open problems currently being pursued in NR research. Since the 90s the field has amassed a great amount of information on the structure and function individual subunits, and in some cases multiple subunits such as DBD-LBD complexes. One of the key challenges moving forward is to integrate this data into the broader context of full-length receptor function and interaction with other members of the transcription machinery. Full-length atomic studies of NRs will be critical to understand allosteric communication within the

receptors, and structural studies of NRs in complex with their interacting partners in the transcription machinery will be necessary to achieve a detailed understanding of transcription in general. Recent technological developments, particularly on the front of cryo-electron microscopy which can now achieve near-atomic resolution of large (>100kD) complexes, should speed progress on this front considerably.

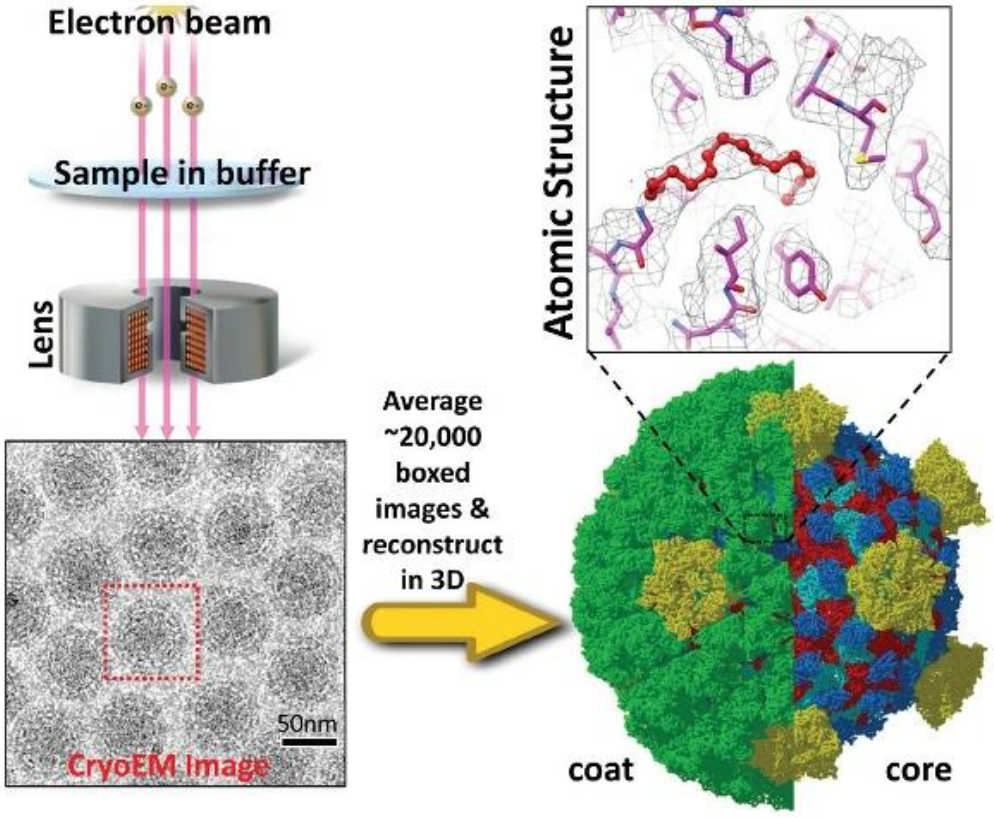

Fig 1.7: Schematic overview of cryo-em as a tool for structural biology.
Source:http://media.americanlaboratory.com/m/20/Article/154443-fig3.jpg

Ultimately, work in this field will move toward a comprehensive understanding of how NRs function within the larger context of transcription regulation, including allosteric mechanisms of communication and interactions with transcription machinery.

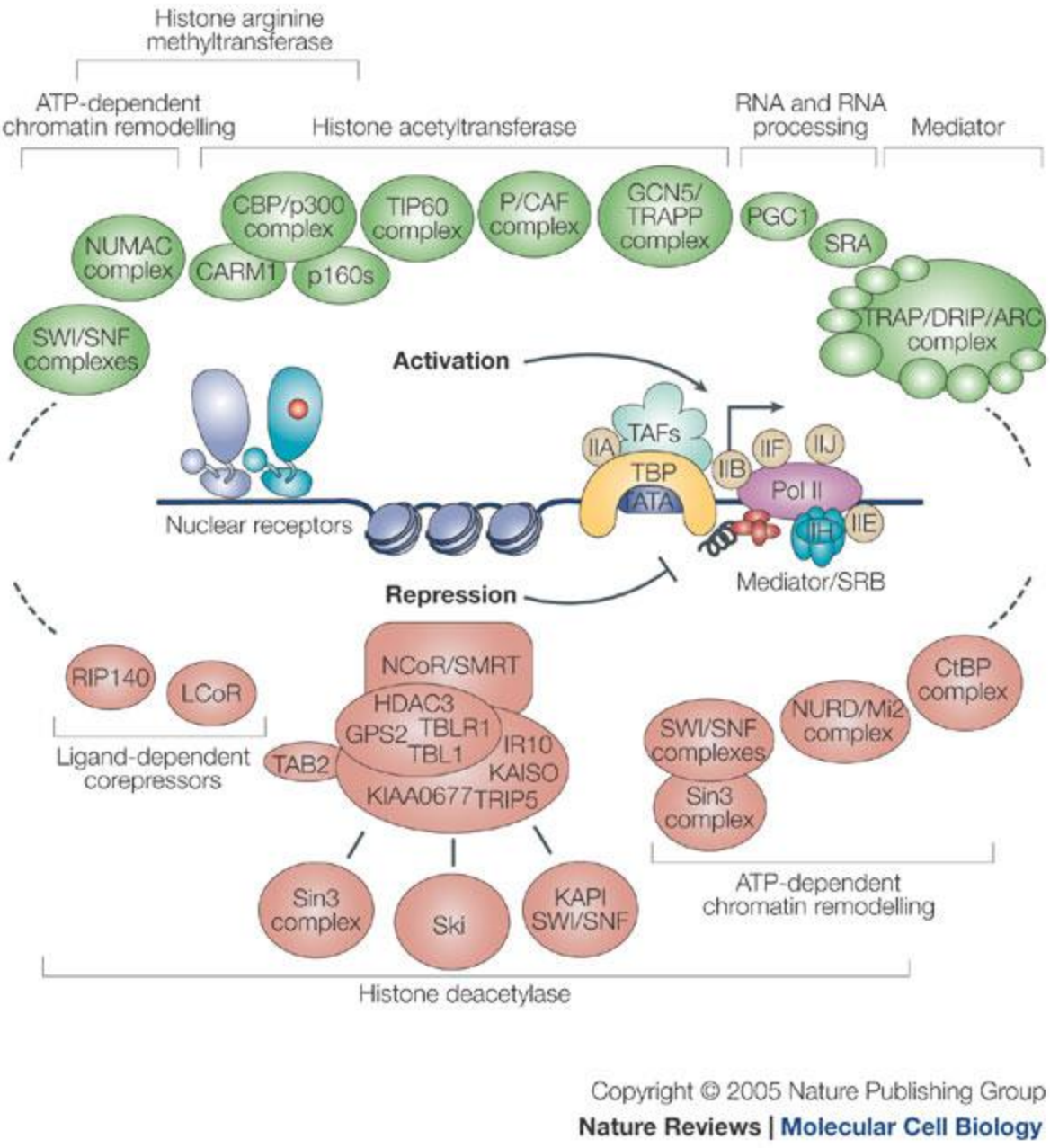

Fig 1.8: Illustration of the many factors involved in the larger transcription initiation machinery. Understanding the interplay of all of these factors will be necessary for a comprehensive understanding of NR transcriptional activity. Source: http://www.nature.com/nrm/journal/v6/n7/images/nrm1680-f2.jpg

This thesis will now briefly diverge from the biology of NRs to describe in detail two biophysical techniques that were essential to the work herein presented. These techniques are Small-angle X-Ray Scattering and Analytical Ultracentrifugation. An overview of the history of the techniques will be presented, followed by a derivation of the basic physics behind the

techniques, concluding with practical considerations for their application to the study of biological molecules such as NRs.

References for Introduction:

- Billas, I.; Moras, D., Allosteric Controls of Nuclear Receptor Function in the Regulation of Transcription. *Journal of Molecular Biology* **2013,** *425* (13), 2317-2329.
- Putcha, B.-D. K.; Wright, E.; Brunzelle, J. S.; Fernandez, E. J., Structural basis for negative cooperativity within agonist-bound TR:RXR heterodimers. *Proceedings of the National Academy of Sciences* **2012,** *109* (16), 6084-6087.
- Rochel, N.; Ciesielski, F.; Godet, J.; Moman, E.; Roessle, M.; Peluso-Iltis, C.; Moulin, M.; Haertlein, M.; Callow, P.; Mély, Y.; Svergun, D. I.; Moras, D., Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. *Nat Struct Mol Biol* **2011,** *18* (5), 564-570.
  Bain, D. L.; Heneghan, A. F.; Connaghan-Jones, K. D.; Miura, M. T., Nuclear Receptor Structure: Implications for Function. *Annual Review of Physiology* **2007,** *69* (1), 201-220.
- Germain, P.; Iyer, J.; Zechel, C.; Gronemeyer, H., Co-regulator recruitment and the mechanism of retinoic acid receptor synergy. *Nature* **2002,** *415* (6868), 187-192.
- Tata, J. R., Signalling through nuclear receptors. *Nat Rev Mol Cell Biol* **2002,** *3* (9), 702-710.
- Li, J.; O'Malley, B. W.; Wong, J., p300 Requires Its Histone Acetyltransferase Activity and SRC-1 Interaction Domain To Facilitate Thyroid Hormone Receptor Activation in Chromatin. *Molecular and Cellular Biology* **2000,** *20* (6), 2031-2042.
- Mulheirn, G., Robinson, Woodward and the synthesis of cholesterol. *Endeavour* **2000,** *24* (3), 107-110.
- Xu, L.; Glass, C. K.; Rosenfeld, M. G., Coactivator and corepressor complexes in nuclear receptor function. *Current Opinion in Genetics & Development* **1999,** *9* (2), 140-147.
- Westin, S.; Kurokawa, R.; Nolte, R. T.; Wisely, G. B.; McInerney, E. M.; Rose, D. W.; Milburn, M. V.; Rosenfeld, M. G.; Glass, C. K., Interactions controlling the assembly of nuclear-receptor heterodimers and co-activators. *Nature* **1998,** *395* (6698), 199-202.
- Willy, P. J.; Mangelsdorf, D. J., Unique requirements for retinoid-dependent transcriptional activation by the orphan receptor LXR. *Genes & Development* **1997,** *11* (3), 289-298.

Physical Review: Small-angle X-ray Scattering

Through the course of the 19th century several different kinds of radiation were observed and characterized, e.g. the discovery of infrared by Hershel in 1800, the discovery of ultraviolet by Ritter in 1801, and the discovery of x-rays by Röntgen in 1895. It was Maxwell's insights into the mathematical description of the electromagnetic field which led to the understanding that these forms of radiation are all disturbances in the electromagnetic field, thereby composing the electromagnetic spectrum.

The understanding of the electromagnetic spectrum allowed for a wide variety of applications by using appropriate wavelength radiation to target specific interactions with matter. X-rays were used in their first medical application only months after Röntgen's discovery, utilizing the fact that Calcium in bones strongly absorbs x-rays and therefore makes bones highly visible in contrast to other biological matter.

As understanding of the nature of x-rays developed, so too did applications. A seminal breakthrough was achieved by Paul Ewald and Max von Laue in 1912, when they postulated that x-rays should have a wavelength conducive to the study of crystal spacings by diffraction. Von Laue experimented with this idea by bombarding a crystal of copper sulfate with x-rays and recording the resulting diffraction. Von Laue latter developed mathematical relations between the scattering angles from the beam to the size and orientation of spacings between unit cells in the crystal. This discovery led to a Nobel Prize in physics in 1914, and ushered in a century of incredible discovery and development with crystallography.

William Bragg and his son together shared the Nobel Prize in Physics in 1915 for their work on crystallographic analysis, which included the development of Bragg's law. These developments paved the way for structural determination of molecules by x-ray crystallography. Simple organic molecules were first solved in the 1920's, followed by larger molecules such as cholesterol and penicillin. Finally, the age of protein crystallography was ushered in by the determination of the structure of myoglobin by Sir John Kendrew in 1958.

Though much of the interest and focus of structural studies was on crystallographic analysis, there was concurrent development on x-ray techniques that do not required highly ordered samples. Guinier and Fournet were pioneers in Small-angle X-ray Scattering (SAXS) studies, discovering in the 1930's that analyzing scattering patterns of molecules in solution

yielded information of the sizes, shapes, and internal structures of the molecules. This technique remained of interest for samples that were impossible to crystallize, but was difficult to apply more generally because of technological limitations. The 1970's brought a revolution in small-angle scattering studies, as the development of synchrotrons provided much higher luminosity beamlines for researchers. Initially, only a few general parameters about the particles, e.g. radius of gyration and volume, were attainable from SAXS curves. The 1990's brought another revolution in SAXS through novel mathematical methods of analysis, which allowed *ab initio* generation of envelope models of the particles to up to 10 angstrom resolution. Currently SAXS is experiencing a strong revival of interest due to the powerful new techniques and software available for structural analysis of scattering curves coupled with improving beamline technology. Though much of the current work in SAXS now focuses on improvements in engineering and application, there continue to be advances in the basic theory, as well.

A succinct overview of x-ray scattering theory will now be presented. A very general diagram of a small-angle scattering experiment is provided:

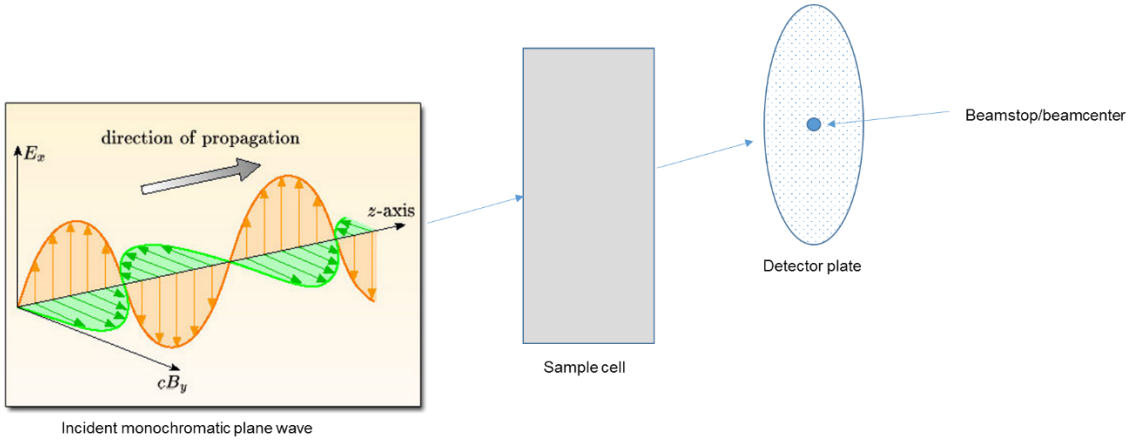

Figure 2.1: Simple diagram of a scattering experiment setup.

A highly collimated plane wave of approximately .1nm x-rays is directed into the sample cell. The electron clouds of the atoms in the sample scatter the waves elastically, and the resulting

waves are recorded on the detector plate. A beamstop is required to block the direct beam, which is many orders of magnitude higher intensity than the scattered radiation.

To understand how reconstruction of particle properties can be obtained from a scattering pattern, it is necessary to understand how x-rays interact with matter. There are several different interactions that incident x-rays may have with matter, and we will classify these as elastic or inelastic and coherent or incoherent. For elastic scattering, the incident plane wave causes oscillation of electrons according to the electric component of the wave, creating an oscillating electric dipole which radiates energy at an equal wavelength in an approximately spherical wave. Inelastically, photons of the incident wave may collide with electrons causing a slight but appreciable rebounding of the electron, which as a result radiates energy of a slightly different wavelength. X-rays may also excite electrons to higher energy levels, which then radiate upon following back down to a lower energy state, but do not in general radiate at the same wavelength as the incident radiation. X-ray interactions with the nucleus are negligible and will not be considered. Finally, a strong enough incident wave may ionize a sample via the photoelectric effect, though this is also a negligible effect in most small-angle scattering studies. We then characterize these interactions as coherent or incoherent, based on whether or not information may be discerned from them. Coherent scattering is Thomson scattering, i.e. radiation caused by dipole oscillation of electrons. Incoherent scattering comprises Compton scattering, where an electron rebounds off of a photon, Absorption, the Photoelectric effect, thermal fluctuations in the sample, and imperfections in sample homogeneity.

Finally, before we begin mathematical analysis of the scattering we must make an approximation known as the Born approximation. Under this approximation we assume that the incident beam is so many orders of magnitude more intense than the weak scattered waves that it

is not appreciable changed as it passes through a sample, and therefore all scattering centers in the sample are subject to the same beam. This approximation holds for most small-angle scattering studies where the particles being studied are very weak scattering sources.

For a scattering experiment we define a differential scattering cross section as well as a solid angle into which the incident radiation through the scattering cross section is scattered. We define the scattering angle as the angle between the scattered and incident beam, and the scattering length density, b, as a measure of the amplitude of a wave observed per unit incident radiation in unit solid angle.

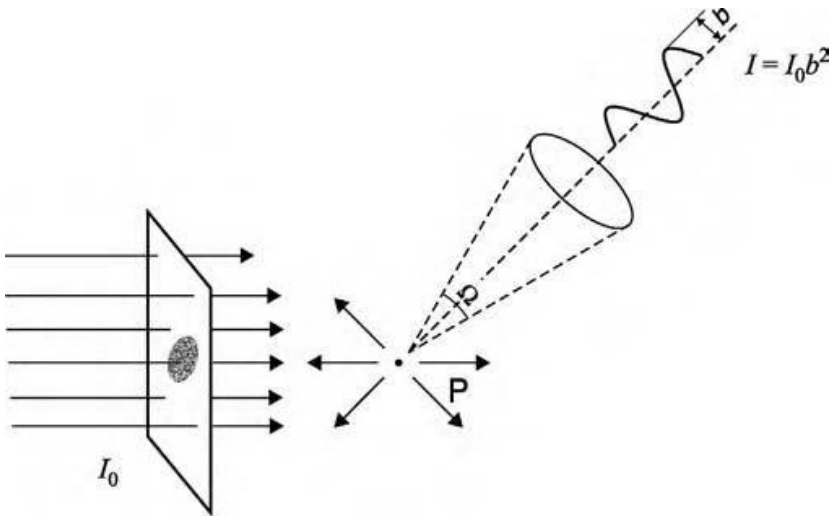

Figure 2.2: Diagram of the mathematical model used to describe scattering. Source: Serdyuk, Zaccai & Zaccai, *Methods in Molecular Biophysics*, Cambridge University Press, 2007.

We require a single arbitrary origin to compare wave phases from different scattering sources, and so with any arbitrary center we have the following diagram:

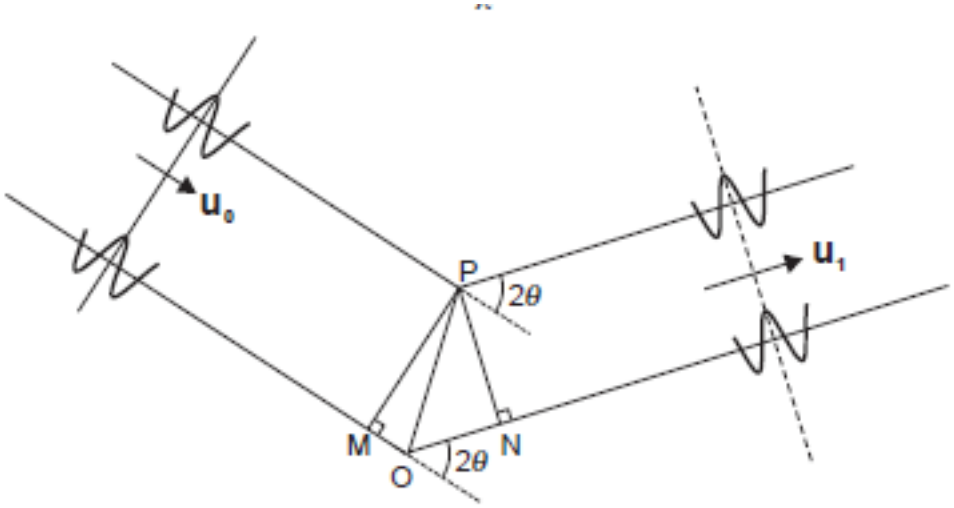

Figure 2.3: Modelling the contribution of distance from an arbitrary origin to the phase of the scattered waves. Source: Serdyuk, Zaccai & Zaccai, *Methods in Molecular Biophysics*, Cambridge University Press, 2007.

Where O is the reference point. This allows us to calculate the phase differences between waves. From the diagram we can see that:

$$\delta = \frac{2\pi(ON - OM)}{\lambda}$$

So the scattered wave relative to O from an atom with scattering amplitude f can be written

$$A = f exp(i\delta)$$

Or, using wavevector notation and defining $Q = k_1 - k_0$

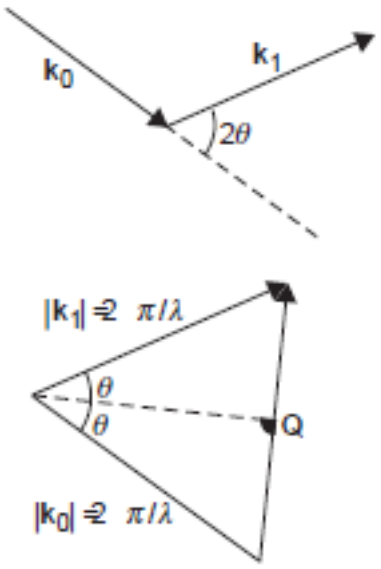

Figure 2.4: Wavevector notation, including the definition of scattering angle and q. Source: Serdyuk, Zaccai & Zaccai, *Methods in Molecular Biophysics*, Cambridge University Press, 2007.

So that we have

$$A = fexp(iq \cdot r) \text{ where } q = \frac{4\pi \sin(\theta)}{\lambda} \text{ from the diagram}$$

Extrapolating to many atoms is simple:

$$F(q) = \sum f_j \exp(iq \cdot r)$$

This begs for a Fourier transform, so we acquiesce

$$f(r) = \int F(q)\exp(-iq \cdot r)\, dV_q \text{ and } F(q) = f(r)\exp \int (iq \cdot r)\, dV_r$$

This would seemingly give us the scattering amplitude as a function of position, from which we could deduce atomic coordinates. However we can only measure $|F(q)|^2$ on the detector.

For a solution of non-interacting, monodisperse and homogeneous particles we have

$$I_n(q) = \sum |F_n(q)|^2$$

But evidently this is not accurate because the particles in solution are in different orientations. To solve this, we approximate that, given enough non-interacting particles, scattering from particles

in different orientations is equivalent to scattering from a particle averaged over all orientations. Therefore we calculate that:

$$< |F(q)|^2 > = \sum_j \sum_k \left(f_j - \rho^0 v_j\right)\left(f_k - \rho^0 v_k\right) \frac{\sin(qr_{jk})}{qr_{jk}} \rightarrow$$

$$I_N(q) = N \sum_j \sum_k m_j m_k \frac{\sin\left(qr_{jk}\right)}{qr_{jk}}$$

Where the terms $f_j - \rho^0 v_j$ represent the contrast amplitude of the jth atom, and m_j is shorthand for this term. We can also write this:

$$I_N(q) = N \iint (\rho(r_j) - \rho^0)(\rho(r_k) - \rho^0) \frac{\sin\left(qr_{jk}\right)}{qr_{jk}} dv_j dv_k$$

Or, normalized by concentration:

$$\frac{I_N(q)}{C} = \frac{N_A}{M} \sum_j \sum_k m_j m_k \frac{\sin\left(qr_{jk}\right)}{qr_{jk}}$$

Where N_A is Avogadro's number, M is the molecular weight, and C is the concentration. This forms the basic foundation for scattering theory.

Now we will look at mathematical developments that allow for intricate analysis of the scattering curve. A major development came from Guinier, who noticed that the low-q part of the curve fits well to a Gaussian:

$$I(q) = I(0) \exp\left[-\frac{1}{3} R_g^2 q^2\right] \rightarrow \ln\left(I(q)\right) = \ln\left(I(0)\right) - \frac{1}{3} R_g^2 q^2$$

This allows for quick and simple determination of I(0) and Rg, as well as data quality validation by analyzing the quality of the fit. The Guinier fit, as it is now described, is one of the most valuable parameters extracted from a scattering curve, and therefore high quality low-q data is very important.

Now we turn to using the entire curve. Taking the inverse Fourier transform of I(q) we get

$$V\gamma(r) = \frac{1}{2\pi^2}\int_0^\infty q^2 I(q)\frac{\sin(qr)}{qr}dq \text{ where } \gamma(r) = < (\rho(r_j) - \rho^0)(\rho(r_k) - \rho^0) >$$

Where V is the volume of a single particle and $\gamma(r)$ is the correlation function. From before we had

$$I_N(q) = N\iint (\rho(r_j) - \rho^0)(\rho(r_k) - \rho^0)\frac{\sin(qr_{jk})}{qr_{jk}}dv_j dv_k$$

So now we can Fourier transform $V\gamma(r)$ to give

$$I(q) = V\int 4\pi r^2\gamma(r)\frac{\sin(qr_{jk})}{qr_{jk}}dr \rightarrow I(q) = \int p(r)\frac{\sin(qr_{jk})}{qr_{jk}}dr$$

Where $p(r) = 4\pi r^2 V\gamma(r)$ is the pairwise distribution function, so that now

$$p(r) = \frac{2}{\pi}r^2\int_0^\infty q^2 I(q)\frac{\sin(qr_{jk})}{qr_{jk}}dq$$

However it is essential to note that this calculation requires values of I(q) at 0 and ∞, both of which must be approximated. Now utilizing this information we can obtain:

$$\int_0^{Dmax} p(r)r^2\,dr = |\int (p(r) - p^0)dv|^2 = |\sum m_j|^2 = I(0)$$

Where Dmax is the maximum dimension of the particle. We also obtain:

$$R_g^2 = \frac{\frac{1}{2}\int_0^{Dmax} p(r)r^2\,dr}{\int_0^{Dmax} p(r)\,dr}$$

And the Porod invariant

$$C = \int_0^\infty I(q)q^2 dq$$

And finally the volume

$$V = \frac{2\pi^2 I(0)}{C}$$

A good test of data quality is comparing the values obtained from I(0) and Rg obtained from the Guinier approximation and integrating the pairwise distribution function. Large discrepancies indicate poor data quality. By using the entire curve one can generate envelopes of particles based on the pairwise distribution of scattering centers, though the mathematical models for *ab inito* envelope reconstruction are beyond the scope of this review.

There are several factors that must be taken into account when preparing for an experiment. For high-quality scattering data one should attempt to isolate a sample that is monodisperse and homogeneous in both composition and conformation. Methods exist for analyzing scattering from heterogeneous systems, but these analyses are more difficult and should therefore be avoided if possible. A concentration series should also be used to extrapolate out any interparticle effects that arise as a consequence of high concentration of particles. High concentration solutions are required for high-q data, but will be most likely to experience inter-particle effects or aggregation and should therefore be complimented with low-concentration samples that can correct for these effects. Analyzing scattering data can be a highly subjective process due to the relative lack of information in a scattering curve as compared to, e.g., a crystal diffraction data set. Therefore, any additional data that may be obtained on the size, geometry, mass, flexibility, and oligomeric state of the particles is very valuable towards constructive SAXS analysis. Finally, radiation damage should be considered as a factor in the experiment. The effect of radiation damage can be minimized by addition of compounds such as glycerol to the buffer, and samples can be analyzed under multiple exposure times to check for radiation damage.

In an ideal experiment, a SAXS dataset on a monodisperse, homogeneous sample with a broad concentration series range and no interparticle effects will return reasonably accurate information on particle size, geometry, radius of gyration, volume, molecular weight, as well as a molecular envelope up to a maximum resolution of 1nm. This information, especially when coupled with data obtained via other methods e.g. crystallography or binding experiments, can yield very valuable structural information on targets that are otherwise hard to analyze structurally. SAXS therefore has a promising future in structural biology for analysis of multimolecular complexes and disordered systems.

Physical Review: Analytical Ultracentrifugation

Analytical ultracentrifugation began its history with Theodore Svedberg's work on gold particles in the 1920's. Svedberg was initially interested in calculating molecular weights of colloidal gold particles using a gravitational sedimentation system coupled with an optical observation system. Svedberg quickly realized that particles of sufficiently small size would require much higher gravitational fields to be able to sediment appreciable, and this would require a very powerful centrifuge. This led to the first design of an analytical centrifuge: a centrifuge capable of achieving very high gravitational fields in the sample cell coupled to an optical system for tracking particle motion. These first designs were largely unsuccessful, however, due to poor sample cell design that allowed for convection to complicate the hydrodynamics of the system. In 1924 Svedberg was able to design a centrifugal system that minimized convection using sector-shaped cells, and it was around this same time that Svedberg's interests shifted from gold colloids to proteins, whose characteristics were still almost entirely unknown at the time. In pioneering equilibrium studies on hemoglobin and

albumin, Svedberg was able to make some essential observations: the molecular weight of the molecules, and the homogeneity of the solutions, which hinted that proteins were large macromolecules that could be identically reproduced in the cell. These observations constituted a great advance in protein science, and yielded Svedberg the Nobel Prize in 1926. By the end of the 1920's, Svedberg and his colleagues had established analytical ultracentrifugation both in theory and instrumentation as an extremely powerful technique for studying biological molecules, though widespread use of the technique remained limited due to high cost and insufficiently advanced instrumentation.

The next great advance in ultracentrifugation came after a massive increase in science funding in the United States following the Second World War. In 1950's a new ultracentrifuge was unveiled: the Spinco model E. This centrifuge was much more versatile and user friendly than the complicated Svedberg centrifuges, and was rapidly adopted by many labs. These centrifuges allowed for a massive increase in the use of centrifugation techniques, and the next two decades saw large accumulations of molecular data from ultracentrifugation techniques, concurrent with advances in instrumentation, methodology, and theory.

The 1970's and the advent of molecular biology techniques such as gel electrophoresis and sequencing led to a very rapid decline in interest in ultracentrifugation. The technique became expensive and difficult in comparison to simpler methods that sufficiently answered the relevant questions of molecular weight and oligomeric status. While advances in theory and data collection and acquisition continued, the use of ultracentrifugation as a technique in biochemistry was rapidly dying. Two decades of incredible progress in ultracentrifugation were followed by two decades of stagnation.

In 1992 Beckman introduced a new ultracentrifuge: the XL-A. The XL-A provided two

essential improvements: a universal set of hardware and rapid digital data acquisition. This centrifuge promised simple and rapid data acquisition and analysis. This led to a sudden explosion of interest in analytical ultracentrifugation, contributing to an impressive rebirth of the technique as one of the gold standards of biochemical and biophysical analysis. Further advances in data analysis continued to improve the accessibility of analytical ultracentrifugation for researchers, and the introduction in 1996 of the Beckman XL-I, which allowed for Raleigh interference optics in addition to the standard absorption optics, provided ultracentrifugation with the versatility required to compete with other techniques to be a valuable tool in the biochemist and biophysicist's arsenal.

Analytical ultracentrifugation has continued to advance in methodology, theory and instrumentation over the years, now boasting a fluorescence optical system in addition to its absorbance and interference systems, and programs to collect and analyze data continue to improve and become more streamlined. Ultracentrifugation is currently taught in many graduate programs as one of the essential tools for biochemical and biophysical studies of molecules, and it is expected that ultracentrifugation will continue to be an invaluable tool for many years to come.

Fundamentally, AUC is a simple technique based on the sedimentation of particles in a solution when exposed to a centrifugal force. Balancing the forces yields equations that depend on the particle mass and geometry and include the particle's motion, therefore by following the motion of a particle (or a boundary) it is possible to deduce molecular parameters of the sample. To describe the motion in a sample cell, we use Fick's diffusion equation with drift included:

$$J_x = -D\frac{dC}{dx} + uC(x)$$

Where D is the diffusion coefficient, u is the angular velocity, C(x) is the concentration as a function of x. But the velocity of the particle is dependent on the angular velocity of the cell, the position of the particle, and a sedimentation coefficient, in other words $u = s\omega^2 x$, so:

$$J_x = -D\frac{dC}{dx} + s\omega^2 xC(x)$$

But we must take into account the geometry of the sample cell. The relevant continuity equation is that

$$\frac{dC}{dt}r = -\frac{1}{r}\frac{drJ}{dr}t$$

So we ultimately obtain the Lamm equation:

$$\frac{dC}{dt}r = -\frac{1}{r}\left[\frac{d}{dr}\left(\omega^2 r^2 sC - Dr\frac{dc}{dr}t\right)\right]t$$

Analytical solutions of the Lamm equation are not attainable except under very specific circumstances that are not particularly useful. However, several numerical methods exist to approximate solutions, though these methods will not be described in detail.

The sedimentation coefficient computed from the Lamm equation solutions is an extremely valuable experimental parameter. To see why, we need to examine how the sedimentation coefficient relates to other parameters. We begin by examining the forces on the sample, of which there are three: the centrifugal, the buoyant, and the frictional.

$$F_c = m\omega^2 r$$

$$F_b = -m_0\omega^2 r$$

$$F_d = -fu$$

Where m, m_0 are the masses of the sample and solvent displaced, r is the distance from rotor center, u is the velocity of the molecule, f is the frictional coefficient, and ω is the angular velocity. Equilibrium is obtained upon balancing forces, which gives

$$m\omega^2 r(1-\nu\rho_0) = fu$$

Where ν is the partial specific volume of the molecule. On a mole basis

$$s \coloneqq \frac{u}{\omega^2 r} = M(1-\nu\rho_0)/N_A f$$

This is the sedimentation coefficient, in units of seconds, and is one of the main values sought after in an AUC experiment.

Modelling the boundary movement in terms of ρ, η, ν, and f/f_0 allows a calculation of the C(s) distribution, which sorts molecules in the sample by sedimentation coefficient, from which can then be derived a C(M) distribution, which sorts molecules in the sample by molecular weight, which is the approach used in SEDFIT.

An important consideration is that the sedimentation coefficient recorded in buffer must be corrected for the standard values expected in water at 20 degrees Celsius, rather than buffered solution at lower temperature.

$$s_{20,w} = s_{exp}\frac{1-\nu\rho_{20,w}}{1-\nu\rho_{exp}}\frac{\eta_{exp}}{\eta_{20,w}}$$

Molecular mass may also be calculated from measurement of the sedimentation and diffusion data, though it is prone to errors in measurement of the relevant values.

$$\frac{s}{D} = \frac{M(1-\nu\rho_0)}{RT}$$

More accurate assessments of mass may be obtained from equilibrium experiments, wherein the sample is centrifuged at a high speed for a long enough time to establish equilibrium of the forces on the sample. This means that the total flux is 0, which means:

$$\frac{dC}{dt}r = -\frac{1}{r}\left[\frac{d}{dr}\left(\omega^2 r^2 sC - Dr\frac{dc}{dr}t\right)\right]t = 0 \rightarrow D(\frac{dC}{dr})_t - \omega^2 rCs = 0$$

Rearranging we get

$$\frac{\mathrm{dln}(C)}{d\frac{r^2}{2}} = \frac{1}{r} C \frac{dC}{dr} \frac{\omega^2 s}{D} = \frac{M(1-\nu\rho)\omega^2}{RT}$$

The concentration distribution falls off exponentially from the meniscus (a) and a point (r):

$$C(r) = C(a) \exp\left[\frac{\omega^2 M(1-\nu\rho)(r^2-a^2)}{2RT}\right]$$

Plotting $ln(C(r))$ vs $r2/2$ yields a straight line for a monodisperse solution, whose slope is $M(1-\nu\rho 0)\omega 2/RT$. For a polydisperse solution, the tangential slope describes gives the average molecular mass at that point in the cell.

These derivations have provided the relevant equations for basic analytical ultracentrifugation experiments running either velocity or equilibrium setups. The programs designed to handle the data acquisition and analysis are described elsewhere.

There are several practical considerations to take into account when setting up an analytical ultracentrifugation experiment. One should first consider which setup, velocity or equilibrium, will be most appropriate for the parameters being studied. Velocity experiments provide information on particle mass, size, shape, and interactions, while equilibrium provides highly accurate data on mass and interactions, and can also be used to test non-ideal behavior of molecules. Next, rotor speeds should be chosen to best fit the system being studied. With prior knowledge of macromolecular composition, optimal rotor speeds can be approximated based on the relevant equations. For example, if either molecular weight or sedimentation coefficient of the macromolecules can be estimated, then the optimal rotor speeds for an equilibrium experiment can be calculated from the final equation provided above. Finally, buffer solutions should be as close to water as feasible in order to minimize error in computing standard values for the experimental parameters. Additions such as glycerol, due to its high viscosity, can strongly contribute to error in the experiment.

References for Physical Reviews:

Textbooks:

- Serdyuk, Zaccai & Zaccai, *Methods in Molecular Biophysics,* Cambridge University Press, 2007.
- Taylor, John R. *Classical Mechanics*. Sausalito, CA: U Science, 2005. Print.
- Griffiths, David J. *Introduction to Quantum Mechanics*. Upper Saddle River, NJ: Pearson Prentice Hall, 2005. Print.
- Sakurai, Jun John., and Jim Napolitano. *Modern Quantum Mechanics*. Boston: Addison-Wesley, 2011. Print.
- Griffiths, David J. *Introduction to Electrodynamics*. Upper Saddle River, NJ: Prentice Hall, 1999. Print.
- Jackson, John David. *Classical Electrodynamics*. New York: Wiley, 1962. Print.

Articles:

- van Holde KE, Hansen JC. Analytical ultracentrifugation from 1924 to the present: A remarkable history. Chemtracts-Biochem Mol Biol. **1998**;11:933–943.
- Jacques, D. A.; Trewhella, J., Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls. *Protein Science* **2010,** *19* (4), 642-657.
- Henderson, R., The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly Reviews of Biophysics* **1995,** *28* (02), 171-193.
- Dmitri, I. S.; Michel, H. J. K., Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics* **2003,** *66* (10), 1735.
- Blanchet, C. E.; Svergun, D. I., Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution. *Annual Review of Physical Chemistry* **2013,** *64* (1), 37-54.

Results:

Agonist-Ligands Mediate the Transcriptional Response of Nuclear Receptor Heterodimers through Distinct Stoichiometric Assemblies with Coactivators

ABSTRACT

The constitutive androstane (CAR) and retinoid X receptors (RXR) are ligand-mediated transcription factors of the nuclear receptor protein superfamily. Functional CAR:RXR heterodimers recruit coactivator proteins, such as the steroid receptor coactivator-1 (SRC1). Here, we show that agonist ligands can potentiate transactivation

through both coactivator binding sites on CAR:RXR which distinctly bind two SRC1 molecules. We also observe that SRC1 transitions from a structurally plastic to a compact form upon binding CAR:RXR. Using small-angle X-ray scattering (SAXS) we show that the CAR(*tcp*):RXR(*9c*)/SRC1 complex can encompass two SRC1 molecules compared to the CAR(*tcp*):RXR/SRC1 which binds only a single SRC1. Moreover, sedimentation coefficients and molecular weights determined by analytical ultracentrifugation (AUC) confirm the SAXS model. Cell-based transcription assays show that disrupting the SRC1 binding site on RXR alters the transactivation by CAR:RXR. These data suggest a broader role for RXR within heterodimers while offering multiple strategies for the assembly of the transcription complex.

Nuclear hormone receptors (NR) relay cellular signals through distinct multiprotein assemblies (1). At the basic level, small-molecule signals produce structural changes within NRs and these changes determine the composition of the interacting protein complex. These changes are essential for transcriptional activity and appear to be conserved among all ligand-activated receptors that have been studied to date. NRs are characteristically modular proteins with distinct functional domains (2). At the N-terminus is the DNA binding domain (DBD) which determines target-gene selectivity. The ligand-binding domain (LBD) is a multi-functional module that contains the ligand-binding pocket, a dimerization interface that associates with the retinoid X receptor (RXR) and a C-terminal ligand-dependent transactivation domain (AF2). Multiple biochemical and structural studies on nuclear receptors have demonstrated that ligand binding results in the specific conformational changes that are associated with a transcriptionally active state (3). In this active state, the conformation of the AF2 domains typically rearrange along the receptor surface, thereby creating a new docking site for transcriptional coactivator proteins (4).

Since both receptors within NR heterodimers can bind small-molecule agonist ligands, in the simplest model for transactivation, agonist binding to either receptor can generate comparable transcriptional levels of downstream genes. Furthermore, this model would predicate that the presence of agonists to both receptors at once would yield proportionately higher levels of transcription. Such model systems are exemplified by the CAR:RXR (Figure 1A), PPARα:RXR and LXR:RXR heterodimers (5). Yet, there are NR heterodimers that exhibit transcriptional responses that are distinct from this model (5,6). For instance, transactivation by RAR:RXR, VDR:RXR and TR:RXR only occurs in the presence of the RAR, VDR and TR agonists, and when used in combination with the RXR agonist, transactivation levels are either enhanced, unaffected or are repressed, respectively (5,7). Through structural and biophysical studies, the mechanism of transactivation has been recognized to occur through conformational changes that restrict recruitment to a single coactivator protein to RAR:RXR (8) or decrease T3 agonist binding-affinity to TR:RXR (7).

CAR is most abundantly expressed in the liver and intestine and has been directly linked to the transcription of genes involved in the clearance of both xenobiotics (9-11), and endogenous toxins such as bilirubin (12). These target genes include select P450 family monooxygenases, phase II conjugating enzymes and xenobiotic transporters. Therefore, CAR serves as a master regulator of xenobiotic clearance and its activation can be considered a form of chemical

immunity. Within the nucleus, CAR binds to RXR and forms a functional heterodimer that recognizes its specific target genes. Additionally, the transcriptional activity of CAR is induced simply by association with RXR and with no apparent need for a CAR ligand (13-15). Although ligand is not required for activation, constitutive CAR activity is mediated through the same conserved functional domains as those utilized by ligand-activated receptors; thus the CAR:RXR heterodimer recruits coactivator proteins through the AF2 transactivation domain (16,17). Transactivation levels mediated by CAR:RXR can be augmented by agonist ligands such as 1,4-*bis*[2-(3,5-dichloropyridyloxy)] benzene (*tcp*) (18) and 6-(4-cholorophenyl)imidazo[2,1-*b*][1,3]thiazole-5-carbaldehyde *O*-(3,4-dichlorobenzyl)oxime (*CITCO*) (19), which are selective for mouse and human CAR, respectively, while 9-*cis* retinoic acid (*9c*) can function as an RXR agonist (20). In both cases, these agonists enhance constitutive activity by stabilizing the constitutive AF2-coactivator interaction. Additionally, transactivation by the CAR:RXR heterodimer can be potentiated by the RXR agonist, *9c* (Figure 1A). *9c* binds the ligand binding pocket of RXR and evokes the canonical NR conformational changes that result in direct interactions with coactivators. The SRC coactivator proteins function to recruit the cellular transcriptional machinery to activated NRs (21,22). SRCs also play an essential role as histone acetyltransferases to acetylate histone proteins and consequently enhance transcriptional activity (23). Thus, there is a direct link between agonist ligand binding, coactivator recruitment and the transcription of downstream genes.

The activity of *permissive* NR heterodimers such as CAR:RXR that is potentiated by ligands to both CAR and RXR raises interesting questions about the precise structural assembly of nuclear factors that promote such transactivation. In this study, we propose that with *permissive* NR heterodimers represented by CAR:RXR, the levels of coactivator recruitment are proportional to the liganded state of the heterodimer. Thus, CAR(*tcp*):RXR/SRC1 is a 1:1 NR:coactivator complex while CAR(*tcp*):RXR(*9c*)/SRC1 exists in 1:2 binding stoichiometry. Moreover, these levels of coactivator recruitment are proportional to transcriptional activity (Figure 1A). These data further suggest that in this subset of NR heterodimers RXR performs a substantial role in regulating transcriptional responses within the cell. Because of the polarity of the LXRE DNA/CAR:RXR complex used in the transactivation assays (Experimental Procedures), CAR occupies the 3' half-site, directly upstream from the luciferase gene. Thus, a major role of the coactivator molecule bound to CAR is to recruit the transcriptional machinery needed for luciferase production.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

CAR:RXR was purified as described earlier (24,25). Briefly, the murine CAR LBD (residues 117−358) was subcloned into the pET15b vector with an N-terminal hexahistidine tag from mCAR cDNA kindly provided by Dr. Barry M. Forman. The human RXRα LBD (residues 225−462) was subcloned into the pACYC184 vector was a kind gift from Dr. Bruce Wisely (Glaxo Smith-Kline,

Inc.). Residues 617-769 of SRC1 (Accession no. Q15788) encompassing three nuclear receptor interacting motifs (RIDs) were subcloned into the pET-SUMO vector. CAR:RXR and SRC1 were separately isolated by affinity chromatography column using Ni-NTA resin (Qiagen Inc.) To prepare various CAR:RXR/SRC complexes, CAR:RXR and SRC(RID 1-3) were mixed in a 1:2 molar ratio, after the addition of ligands, and loaded onto an S200 Superdex 16/60 column for purification of the resulting complexes (Figure E1). Fractions corresponding to the complexes were pooled, measured by Bradford Assay, and concentrated for further analysis.

Analytical Ultracentrifugation

Aliquots of CAR:RXR and SRC RID 1-3 were thawed and mixed in various ratios (1:3, 1:1, 3:1), incubated briefly with ligands, and analyzed on a Beckman XL ultracentrifuge. Both sedimentation velocity and sedimentation equilibrium experiments were performed. Sedimentation equilibrium was performed at 8-10ºC and 3 rotor speeds, while sedimentation velocity was run at 20ºC and 55,000rpm. Data was analyzed using SEDFIT and SEDPHAT (26). We calculate the parameters, f/f_0, frictional ratio; sw(20,w), sedimentation coefficient under standard conditions; and *rmsd* which reports the quality fit of the solutions of the Lamm equations (26) to the data.

Small-angle X-ray Scattering

Measurements were recorded at several beamlines: SIBLYS at Laurence Berkeley National Laboratory, DND-CAT at Argonne National Laboratory, and MACCHESS at Cornell. Data was analyzed using the ATSAS software package (27) and ScÅtter. 3D model building was performed using DAMMIN and MONSA (28) and visualized using Chimera. Kratky plots were calculated for shape analysis (29,30). The theoretical values for spherical objects in the V_c-based Kratky plot attain an ordinate maxima of 0.82, and for non-flexible scattering particles, $q^2 * V_c = \sqrt{3}$, where q is the scattering vector (Å) and V_c is defined by the correlation length of the scattering particle as the ratio of the zero angle scattering intensity, $I(0)$, to its total scattered intensity (30).

Reporter gene assays

These assays were performed as reported earlier (31). Briefly, CV-1 cells were maintained in DMEM/F-12 media containing 10% fetal bovine serum and 1000 U/ml penicillin and 1 mg/ml streptomycin. Immediately prior to the assay, the media was changed to DMEM/F-12 with 10% charcoal-dextran treated FBS and no antibiotics. Lipofectamine® (Invitrogen) was used to transfect cells with 50 ng/well pCMX mCAR, 40ng/well pCMX-Gal4-RXR, 100 ng/well pCMV-TK-luc containing three copies of the liver X receptor response element (LXRE) and 10 ng/well of pRL CMV expressing renilla luciferase as an internal control. The cells were dispensed on 24-well plates and ligands were added 24 hours post transfection. The ligand concentrations used were TCPOBOP (1.0 μM), 0.5□M 9-cis retinoic acid. After 48 hours, cells were lysed. Activity was determined using the dual luciferase assay kit (Promega Inc.) following the manufacturer's instructions. The reported results are the average from three separate experiments.

Gal4-DBD assays: The E456K mutation was made on pCMX-Gal4DBD-hRXRαLBD (gift from Prof. Barry Forman). Transfection and assay was performed as described above with full-length CAR and four copies of a Gal4 binding site (pUC8-MH100x4-*TK-Luc*).

RESULTS

CAR(*tcp*):RXR(*9c*)/SRC1 Assembles as a 1:2 Heterodimer:Coactivator Complex

To develop an understanding of how the transcriptionally active CAR:RXR assembles with coactivators, we prepared CAR:RXR LBD and SRC1(RID1-3) in *E. coli* and isolated multiple complexes of CAR:RXR/SRC1 (Supplementary Figure 1). Since CAR:RXR can also bind SRC1 in the absence of agonist, for the study here we isolated the complexes CAR:RXR/SRC1, CAR(*tcp*):RXR/SRC1 and CAR(*tcp*):RXR(*9c*)/SRC1 representing the unliganded, singly-liganded and doubly-liganded CAR:RXR complexes, respectively. SRC1(RID1-3) used in these studies comprises amino acids 617-769, and encompasses the three nuclear receptor interacting domains (RIDs) (Supplementary Figure 1A). Also, SRC1(RID1-3) has previously been shown to interact in a ligand-dependent manner with CAR:RXR (15,25). Using small-angle X-ray scattering (SAXS), we have determined the global molecular assembly and structural properties of the CAR:RXR/SRC1, CAR(*tcp*):RXR/SRC1 and CAR(*tcp*):RXR(*9c*)/SRC1 complexes (Figure 1 and Supplementary Table 1). For comparisons, we also measured scattering from CAR(*tcp*):RXR(*9c*) and SRC1 alone and the shapes of the scattering curves are distinct and typical of their molecular size and flexibility. Therefore, the scattering curve of CAR:RXR heterodimer alone is characteristic of folded protein, while the scattering curve of SRC alone is representative of disordered proteins (Figure 1B). We applied Kratky analyses (29) and shape comparisons to identify any noticeable difference in compactness between CAR(*tcp*):RXR/SRC1 and CAR(*tcp*):RXR(*9c*)/SRC1. From these analyses we could conclude that the CAR(*tcp*):RXR(*9c*)/SRC1 complex is more elongated and shows greater flexibility than CAR(*tcp*):RXR/SRC1, which suggests the presence of a second SRC1 molecule within the CAR(*tcp*):RXR(*9c*)/SRC1 particle. We also compare the $I(0)$-scaled and V_c-based Kratky scattering curves which emphasize differences in size and geometry of the scattering particles (Materials and Methods) (30). In the $I(0)$-normalized graphical plot, we compare the linearity and negative slope of the scaled intensity versus scattering angle for the complexes (Figure 1B). The $I(0)$-normalized scattering plot of CAR(*tcp*):RXR(*9c*)/SRC1 complex is more linear with a sharper slope than the corresponding plots of either CAR(tcp):RXR/SRC1 or CAR:RXR/SRC1 complexes. This clearly suggests that CAR(*tcp*):RXR(*9c*)/SRC1 is a relatively more extended molecule of higher molecular weight than the unliganded and singly-liganded complexes. From the the V_c-Kratky plot using experimental SAXS data we are able to infer that CAR:RXR alone is mostly spherical with no apparent flexibility, while free SRC1 shows a hyperbolic plateau that is indicative of a highly flexible structure and with higher surface area-to-volume ratio (Figure 1C). When comparing the heterodimer:coactivator complexes we note that there is a decrease in peak height between CAR(*tcp*):RXR/SRC1 and CAR(*tcp*):RXR(*9c*)/SRC1 which further indicates that

CAR(*tcp*):RXR(*9c*)/SRC1 is relatively extended particle (Figure 1C and Figure E1E). Overall, the observed scattering pattern is consistent with the presence of a second SRC molecule in the CAR(*tcp*):RXR(*9c*)/SRC complex. Furthermore, we note from a comparison of molecular envelopes of free and bound SRC1 that this molecule adopts a relatively more compact structure upon binding CAR:RXR, a feature that has been observed previously in the RAR:RXR/SRC1 complex (8).

To visualize the assembly of these complexes we generated molecular envelopes of CAR:RXR/SRC1, CAR(*tcp*):RXR/SRC1 and CAR(*tcp*):RXR(*9c*)/SRC1 (Figure 1D, E & F). The molecular shapes of CAR:RXR/SRC1 and CAR(*tcp*):RXR/SRC1 are of an elongated species and are nearly identical at SAXS resolution (Figure 1E). The molecular shape of the CAR(*tcp*):RXR(*9c*)/SRC1 complex is also elongated and is approximately 1.4-fold larger than the CAR:RXR/SRC1 and CAR(*tcp*):RXR/SRC1 complexes (Figure 1E & F and Supplementary Figure 1E). To depict the assembly of each complex, the molecular envelopes of the heterodimer and coactivator were superimposed upon the envelope of each CAR:RXR/SRC1 complex. Both manual and automated fitting (see Material and Methods) suggest that both the CAR:RXR/SRC1, CAR(*tcp*):RXR/SRC1 envelopes can encompass the CAR:RXR heterodimer bound to a single SRC1 molecule (Figure 1E). On the other hand, the molecular shape of CAR(*tcp*):RXR(*9c*)/SRC1 readily corresponds to a single CAR:RXR heterodimer that is bound to two SRC1(RID1-3) molecules (Figure 1F and Supplementary Figure 2). Thus, these SAXS analyses suggest molecular complexes of stoichiometries of 1:1 heterodimer:coactivator for the singly-liganded CAR(*tcp*):RXR/SRC1 and 1:2 heterodimer:coactivator for and CAR(*tcp*):RXR(*9c*)/SRC complexes.

Analytical Ultracentrifugation for Size Determination. To establish if the shapes of CAR:RXR/SRC1, CAR(*tcp*):RXR/SRC1 and CAR(*tcp*):RXR(*9c*)/SRC1 correspond to their relative sizes, these complexes were independently analyzed through sedimentation velocity and equilibrium studies by analytical ultracentrifugation (AUC). The sedimentation velocity data consistently shows a species of higher sedimentation coefficient with CAR(*tcp*):RXR(*9c*)/SRC1 than CAR(*tcp*):RXR/SRC1 (Figure 2A and Table 1). From sedimentation equilibrium analyses we confirmed the molecular weights of these species to be 84.0 *k*Da (CAR:RXR/SRC1 and CAR(*tcp*):RXR/SRC1) which corresponds to the size of one CAR:RXR heterodimer (55*k*Da) bound to a single SRC1 (30*k*Da) and 113.0 *k*Da for the CAR(*tcp*):RXR(*9c*)/SRC1 complex, which corresponds to one CAR:RXR heterodimer bound to two SRC1 molecules (Figure 2A and Supplementary Figure 3). The equilibrium data therefore suggests a strong preference for 2:1 complex formation in the doubly-liganded state. Taken together, the ultracentrifugation data confirm that the molecular envelopes determined by SAXS correspond to the molecular weights of these complexes as determined by AUC. Therefore, we hypothesize that transactivation by CAR(*tcp*):RXR(*9c*)/SRC1 is relies on coactivator binding to both CAR and RXR.

Transactivation by CAR:RXR^{E456K} is Distorted from the Native CAR:RXR Complex

There are multiple factors within the cell that function to regulate transactivation by NRs (32,33). Of these factors, the SRC family of coactivator proteins are recruited specifically to the agonist-bound conformation of the NR LBD (34). In this agonist-bound conformation the AF2 domain is realigned along the receptor surface, and in doing so creates a new interface that can bind SRC proteins (35). The amino acid E456 is within the AF2 domain of RXR and interacts with SRC1 (16). To better understand the role of the RXR coactivator binding site within CAR:RXR, we compare the transactivation of CAR:RXR^{E456K} with the native protein complex in a cell-based reporter system. This E→K amino acid substitution has previously been shown to disrupt SRC1 recruitment by RXR (36). When tested in CV1 cells, we notice that the transcriptional response of CAR:RXR and CAR:RXR^{E456K} to *tcp* (relative to no exogenous ligand) is similar with no noticeable allosteric effects of the RXR^{E456K} mutation on CAR. However, transactivational levels in response to exogenously applied *9c* alone (Figure 2B) and to the combination of *9c+tcp* by the native CAR:RXR and mutant CAR:RXR^{E456K} receptor complexes are markedly different (Supplementary Figure 4). The measurements of SAXS and AUC above propose a distinct role for RXR within the CAR:RXR complex. Together, these results suggest that in CAR:RXR the SRC-binding site within RXR is essential for the heterodimer to achieve maximum transcriptional activity. Thus, relative to other NR heterodimers such as RAR:RXR and VDR:RXR (8,37), RXR can undertake a more significant role in transactivation by CAR:RXR.

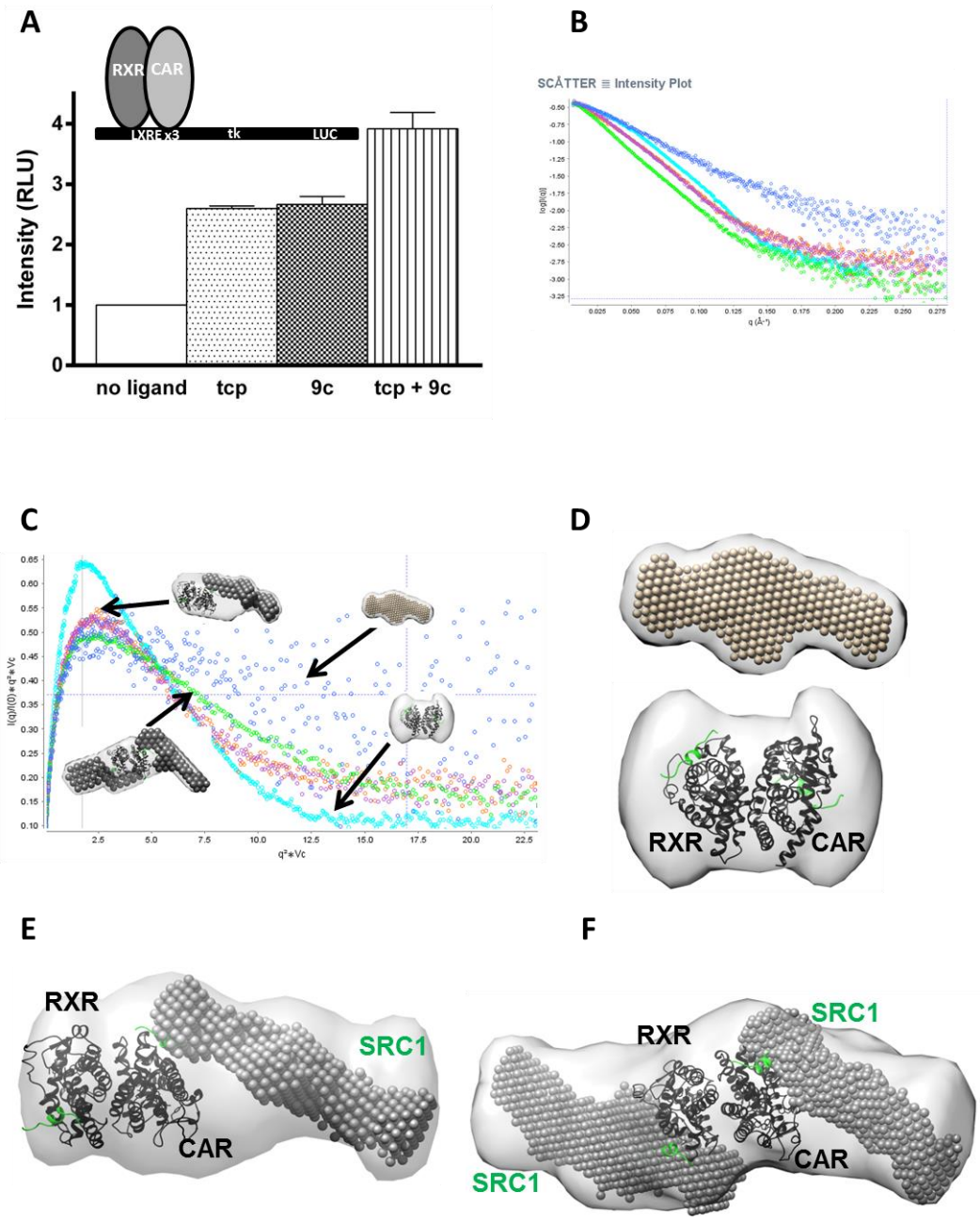

Figure 1. Activity and assembly of CAR:RXR. (A). Transactivation of CAR:RXR measured in CV-1 cells on 3 copies of LXRE. (B) Scattering curve normalized to I(0) to show differences in size and deviation from globular shape. Blue, SRC1; cyan, CAR:RXR; Orange, CAR:RXR/SRC1, purple, CAR(*tcp*):RXR/SRC1; green, CAR(*tcp*):RXR(*9c*)/SRC1. (C) V_c based Kratky plot for visualization of flexibility and surface area-to-volume ratio. Molecular shapes of the complexes are generated by MONSA, while the individual heterodimer and coactivator envelopes are generated by DAMMIN. (D) Top, molecular shape of free SRC1(RID1-3); bottom, molecular shape of CAR(*tcp*):RXR(*9c*). (E) & (F) Molecular shapes of CAR(*tcp*):RXR/SRC and CAR(*tcp*):RXR(*9c*)/SRC (light grey), respectively. Superimposed on this is the SRC (dark grey spheres) envelope from (D) and the CAR:RXR/SRC(peptide, green) structure (PDB ID:1XLS). Molecular envelopes are produced by DAMMIN from SAXS data.

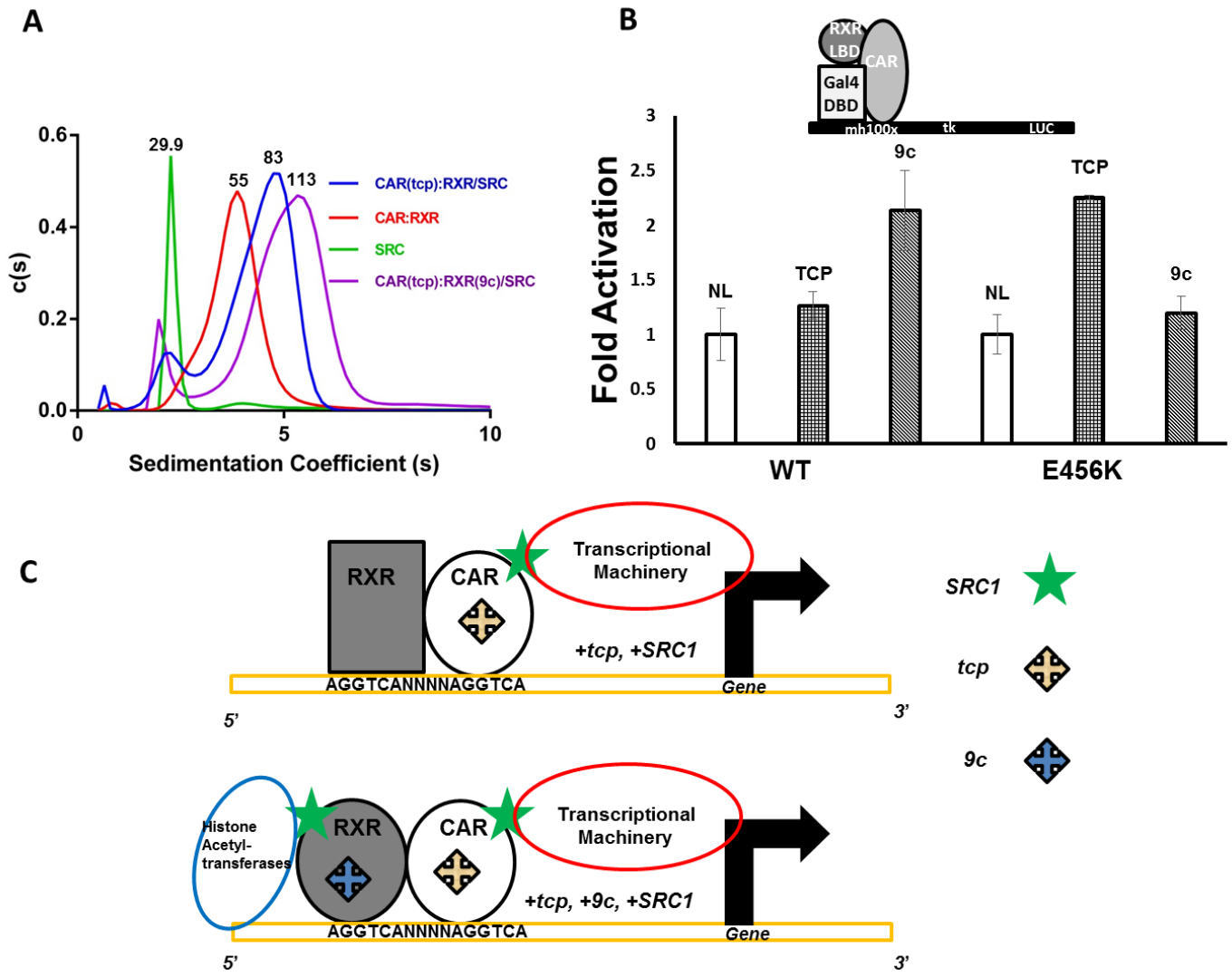

Figure 2. Activity of CAR and SRC1 mutants. (A). Sedimentation coefficients from sedimentation velocity experiments using AUC. Molecular weights shown above each peak are in *k*Da (See also Table 1 and Figure S2). (B). Transactivation by CAR:Gal4 DBD-RXR LBD and CAR:Gal4 DBD-RXR^{E456K} LBD measured in CV-1 cells on four copies of a Gal4 binding site (mh100x4-*tk-luc*) response element. C. Models of the NR transactivation complexes. Top, CAR:RXR/SRC1, CAR(*tcp*):RXR/SRC1. Bottom, CAR(*tcp*):RXR(*9c*)/SRC1.

Molecular Species	f/f_o	sw(20,w)	*rmsd*
CAR:RXR	1.28	4.12	0.005
SRC1	1.5	2.1	0.007
CAR:RXR/SRC	1.46	2.27, 4.7	0.007
CAR(*tcp*):RXR/SRC	1.46	2.4, 4.7	0.006
CAR:RXR(*9c*)/SRC	1.5	2.48, 5.69	0.006
CAR(*tcp*):RXR(*9c*)/SRC	1.5	2.27, 5.6	0.009

Table 1: **AUC velocity data**. f/f_0, frictional ratio; sw(20,w), sedimentation coefficient under standard conditions; *rmsd* reports the quality fit to the data. A higher sedimentation coefficient for *9c* and *tcp+9c* complexes indicates a larger species, while a slightly higher frictional coefficient indicates a more extended structure.

DISCUSSION

Nuclear receptors are a superfamily of structurally and functionally conserved proteins that have evolved to regulate transcription in response to small-molecule ligands through multiprotein assemblies. Hormonal agonist molecules evoke the correct structural changes within NRs to interact directly with coactivators, such as SRC1. **Normal** transactivation is dependent on the precise assembly of the component molecules. However, there is only a superficial mechanistic understanding of how this multiprotein assembly takes place, how it can be modulated and how it relates to transactivation. The study here illustrates the role of the agonist ligand in defining the molecular assembly of the NR:coactivator complex (Figure 2C).

Both, CAR and RXR in the CAR:RXR heterodimer can independently bind their respective agonists (24). Also, crystallographic studies have shown that CAR(*agonist*):RXR(*agonist*) can bind two 13-mer LXXLL coactivator-derived peptides through binding sites on both CAR and RXR (16,17). This structural assembly has also been observed in other *permissive* NR heterodimers such as *LXR:RXR (38) and* PPAR□:RXR (39). It is now clear from our data that these LXXLL motifs that are bound to permissive NR heterodimers are derived from two independent SRC1 molecules, although the intact SRC1 molecule has three distinct LXXLL-containing RIDs (40).

This agonist-mediated heterodimer:coactivator stoichiometry has important mechanistic implications for transactivation and in pharmacology. First, among the several functions ascribed to SRCs are the recruitment of the cellular transcriptional machinery to activated NRs (21,22,41) and as histone acetyltransferases (23). The polarity of heterodimers such as CAR:RXR on the direct repeat response element places CAR towards the 3' end of the promoter (42-44). Therefore, the most likely function of the coactivator molecule bound directly to CAR is to assemble the transcriptional machinery through interactions with p300/CBP (22). The addition of *9c* to permissive NR heterodimers allows for the recruitment of a second SRC coactivator molecule directly to RXR on the 5' end of the promoter. From this location, the second SRC can function as and recruit other histone acetyltransferases thereby enhancing transactivational levels, as observed with the progesterone and glucocorticoid receptors (45) (Figure 2C). Second, this mechanism of transactivation is a distinct alternative to that proposed for the RAR:RXR

heterodimer (8,37,46). In RAR:RXR, a single coactivator molecule is recruited directly to RAR upon activation by agonist in a conformation similar to CAR(*tcp*):RXR/SRC. However, the recruitment of a second coactivator to RAR(*agonist*):RXR(*agonist*) is restricted through long-range, agonist-induced conformational changes that disrupt the RXR coactivator-binding site (47). Third, targeting coactivators for therapy is of growing interest (48,49), thus requiring a detailed knowledge of such binding events. These coactivators display different binding specificities for the receptors both independently (50) and within the heterodimer (15), thus, it is likely that the specific interactions between each SRC1 molecule and the two binding sites on CAR:RXR are distinct. The activation of CAR:RXR is not always beneficial as hepatic metabolism can convert certain therapeutic drugs to potent toxins. For instance, CAR:RXR-mediated metabolism of acetaminophen results in a reactive quinone metabolite (*N*-acetyl-*p*-benzoquinone imine). This metabolic by-product promotes acute liver failure by binding to cellular macromolecules and by generating reactive oxygen species (10,51). The hepatotoxic effects of cocaine are also mediated via a CAR:RXR–dependent pathway (9). Thus, the activity of CAR:RXR can have either protective or deleterious consequences to the organism depending on the particular chemical challenges faced by it. Also, the discoveries of endogenous RXR ligands such as polyunsaturated fatty acids (52-54) increase the likelihood of two agonists binding the CAR:RXR heterodimer at once. Thus, this novel assembly has important implications for the design of small molecules directed at regulating transactivation by modulating the formation and composition of the NR-coactivator assembly.

REFERENCES

1. Rosenfeld, M. G., and Glass, C. K. (2001) *J Biol Chem* 276, 36865-36868
2. Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., and Evans, R. M. (1995) *Cell* 83, 835-839
3. Billas, I., and Moras, D. (2013) *J Mol Biol* 425, 2317-2329
4. Onate, S. A., Boonyaratanakornkit, V., Spencer, T. E., Tsai, S. Y., Tsai, M. J., Edwards, D. P., and O'Malley, B. W. (1998) *J Biol Chem* 273, 12101-12108
5. Shulman, A. I., Larson, C., Mangelsdorf, D. J., and Ranganathan, R. (2004) *Cell* 116, 417-429
6. Forman, B. M., Umesono, K., Chen, J., and Evans, R. M. (1995) *Cell* 81, 541-550
7. Putcha, B. D., Wright, E., Brunzelle, J. S., and Fernandez, E. J. (2012) *Proc Natl Acad Sci U S A* 109, 6084-6087
8. Rochel, N., Ciesielski, F., Godet, J., Moman, E., Roessle, M., Peluso-Iltis, C., Moulin, M., Haertlein, M., Callow, P., Mely, Y., Svergun, D. I., and Moras, D. (2011) *Nat Struct Mol Biol* 18, 564-570
9. Wei, P., Zhang, J., Egan-Hafley, M., Liang, S., and Moore, D. D. (2000) *Nature* 407, 920-923.
10. Zhang, J., Huang, W., Chua, S. S., Wei, P., and Moore, D. D. (2002) *Science* 298, 422-424.
11. Sonoda, J., Rosenfeld, J. M., Xu, L., Evans, R. M., and Xie, W. (2003) *Curr Drug Metab* 4, 59-72
12. Huang, W., Zhang, J., Chua, S. S., Qatanani, M., Han, Y., Granata, R., and Moore, D. D. (2003) *Proc Natl Acad Sci U S A* 100, 4156-4161
13. Baes, M., Gulick, T., Choi, H. S., Martinoli, M. G., Simha, D., and Moore, D. D. (1994) *Mol Cell Biol* 14, 1544-1551.
14. Choi, H. S., Chung, M., Tzameli, I., Simha, D., Lee, Y. K., Seol, W., and Moore, D. D. (1997) *J Biol Chem* 272, 23565-23571.
15. Dussault, I., Lin, M., Hollister, K., Fan, M., Termini, J., Sherman, M. A., and Forman, B. M. (2002) *Mol Cell Biol* 22, 5270-5280
16. Suino, K., Peng, L., Reynolds, R., Li, Y., Cha, J. Y., Repa, J. J., Kliewer, S. A., and Xu, H. E. (2004) *Mol Cell* 16, 893-905
17. Xu, R. X., Lambert, M. H., Wisely, B. B., Warren, E. N., Weinert, E. E., Waitt, G. M., Williams, J. D., Collins, J. L., Moore, L. B., Willson, T. M., and Moore, J. T. (2004) *Mol Cell* 16, 919-928
18. Tzameli, I., Pissios, P., Schuetz, E. G., and Moore, D. D. (2000) *Mol Cell Biol* 20, 2951-2958.
19. Maglich, J. M., Parks, D. J., Moore, L. B., Collins, J. L., Goodwin, B., Billin, A. N., Stoltz, C. A., Kliewer, S. A., Lambert, M. H., Willson, T. M., and Moore, J. T. (2003) *J Biol Chem* 278, 17277-17283
20. Levin, A. A., Sturzenbecker, L. J., Kazmer, S., Bosakowski, T., Huselton, C., Allenby, G., Speck, J., Kratzeisen, C., Rosenberger, M., Lovey, A., and et al. (1992) *Nature* 355, 359-361

21. Kamei, Y., Xu, L., Heinzel, T., Torchia, J., Kurokawa, R., Gloss, B., Lin, S. C., Heyman, R. A., Rose, D. W., Glass, C. K., and Rosenfeld, M. G. (1996) *Cell* 85, 403-414

22. Yao, T. P., Ku, G., Zhou, N., Scully, R., and Livingston, D. M. (1996) *Proc Natl Acad Sci U S A* 93, 10626-10631

23. Spencer, T. E., Jenster, G., Burcin, M. M., Allis, C. D., Zhou, J., Mizzen, C. A., McKenna, N. J., Onate, S. A., Tsai, S. Y., Tsai, M. J., and O'Malley, B. W. (1997) *Nature* 389, 194-198

24. Wright, E., Vincent, J., and Fernandez, E. J. (2007) *Biochemistry* 46, 862-870

25. Wright, E., Busby, S. A., Wisecarver, S., Vincent, J., Griffin, P. R., and Fernandez, E. J. (2011) *Structure* 19, 37-44

26. Schuck, P. (2000) *Biophysical journal* 78, 1606-1619

27. Konarev, P. V., Petoukhov, M. V., Volkov, V. V., and Svergun, D. I. (2006) *J Appl Crystallogr* 39, 277-286

28. Svergun, D. I. (1999) *Biophys J* 76, 2879-2886

29. Receveur-Brechot, V., and Durand, D. (2012) *Current protein & peptide science* 13, 55-75

30. Rambo, R. P., and Tainer, J. A. (2013) *Nature* 496, 477-481

31. Shan, L., Vincent, J., Brunzelle, J. S., Dussault, I., Lin, M., Ianculescu, I., Sherman, M. A., Forman, B. M., and Fernandez, E. J. (2004) *Mol Cell* 16, 907-917

32. Lonard, D. M., and O'Malley, B. W. (2012) *Nature reviews. Endocrinology* 8, 598-604

33. Millard, C. J., Watson, P. J., Fairall, L., and Schwabe, J. W. (2013) *J Mol Endocrinol* 51, T23-36

34. Johnson, A. B., and O'Malley, B. W. (2012) *Mol Cell Endocrinol* 348, 430-439

35. Wurtz, J. M., Bourguet, W., Renaud, J. P., Vivat, V., Chambon, P., Moras, D., and Gronemeyer, H. (1996) *Nat Struct Biol* 3, 87-94

36. Schulman, I. G., Chakravarti, D., Juguilon, H., Romo, A., and Evans, R. M. (1995) *Proc Natl Acad Sci U S A* 92, 8288-8292

37. Germain, P., Iyer, J., Zechel, C., and Gronemeyer, H. (2002) *Nature* 415, 187-192

38. Lou, X., Toresson, G., Benod, C., Suh, J. H., Philips, K. J., Webb, P., and Gustafsson, J. A. (2014) *Nat Struct Mol Biol* 21, 277-281

39. Chandra, V., Huang, P., Hamuro, Y., Raghuram, S., Wang, Y., Burris, T. P., and Rastinejad, F. (2008) *Nature*

40. Heery, D. M., Kalkhoven, E., Hoare, S., and Parker, M. G. (1997) *Nature* 387, 733-736.

41. Kalkhoven, E., Valentine, J. E., Heery, D. M., and Parker, M. G. (1998) *Embo J* 17, 232-243

42. Perlmann, T., Rangarajan, P. N., Umesono, K., and Evans, R. M. (1993) *Genes Dev* 7, 1411-1422

43. Kurokawa, R., Yu, V. C., Naar, A., Kyakumoto, S., Han, Z., Silverman, S., Rosenfeld, M. G., and Glass, C. K. (1993) *Genes Dev* 7, 1423-1435

44. Zechel, C., Shen, X. Q., Chen, J. Y., Chen, Z. P., Chambon, P., and Gronemeyer, H. (1994) *Embo J* 13, 1425-1433

45. Li, X., Wong, J., Tsai, S. Y., Tsai, M. J., and O'Malley, B. W. (2003) *Mol Cell Biol* 23, 3763-3773

46. Westin, S., Kurokawa, R., Nolte, R. T., Wisely, G. B., McInerney, E. M., Rose, D. W., Milburn, M. V., Rosenfeld, M. G., and Glass, C. K. (1998) *Nature* 395, 199-202.

47. Osz, J., Brélivet, Y., Peluso-Iltis, C., Cura, V., Eiler, S., Ruff, M., Bourguet, W., Rochel, N., and Moras, D. (2012) *Proceedings of the National Academy of Sciences* 109, E588-E594

48. Wu, Z., and Boss, O. (2007) *Expert Opin Ther Targets* 11, 1329-1338

49. Hwang, J. Y., Attia, R. R., Zhu, F., Yang, L., Lemoff, A., Jeffries, C., Connelly, M. C., and Guy, R. K. (2012) *J Med Chem* 55, 2301-2310

50. Ding, X. F., Anderson, C. M., Ma, H., Hong, H., Uht, R. M., Kushner, P. J., and Stallcup, M. R. (1998) *Mol Endocrinol* 12, 302-313

51. Gill, R. Q., and Sterling, R. K. (2001) *J Clin Gastroenterol* 33, 191-198

52. de Urquiza, A. M., Liu, S., Sjoberg, M., Zetterstrom, R. H., Griffiths, W., Sjovall, J., and Perlmann, T. (2000) *Science* 290, 2140-2144

53. Lengqvist, J., Mata De Urquiza, A., Bergman, A. C., Willson, T. M., Sjovall, J., Perlmann, T., and Griffiths, W. J. (2004) *Mol Cell Proteomics* 3, 692-703

54. Volakakis, N., Joodmardi, E., and Perlmann, T. (2009) *Biochem Biophys Res Commun* 390, 1186-1191